Business Alignment: Using Process Mining as a Tool for Delta Analysis

W.M.P. van der Aalst

Department of Technology Management, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands. w.m.p.v.d.aalst@tm.tue.nl

Abstract. Fueled by the omnipresence of event logs in transactional information systems (cf. WFM, ERP, CRM, SCM, and B2B systems), process mining has become a vivid research area. Until recently, the information in these event logs was rarely used to analyze the underlying processes. Process mining aims at improving this by providing techniques and tools for discovering process, control, data, organizational, and social structures from event logs, i.e., the basic idea of process mining is to diagnose business processes by mining event logs for knowledge. In this position paper we focus on the potential use of process mining for measuring business alignment, i.e., comparing the real behavior of an information system or its users with the intended or expected behavior. Such a Delta analysis may assist in creating and maintaining the fit between business processes and supporting information systems.

1 Introduction

In many organizations new processes are emerging and existing processes are changing ("The only contant is change"). Therefore, the *alignment of business processes and information systems* requires continuous attention. To "maintain the fit" it is important to detect changes over time, i.e., deviations of the described or prescribed behavior. Rare deviations will always happen, but should not be regarded as a symptom of a change. However, if considerable amount of process instances deviate from the prescribed pattern, some action should be undertaken to align the business process and the supporting information system. Similarly, to "create the fit" it may be worthwhile to monitor the actual behavior of the people in the organization before configuring/installing the (new) information system. Both for creating and maintaining the fit (i.e., securing alignment) we propose *process mining* techniques which use event logs to discover the actual process.

Today, many enterprise information systems store relevant events in some structured form. For example, workflow management systems typically register the start and completion of activities. ERP systems like SAP log all transactions, e.g., users filling out forms, changing documents, etc. Business-to-business (B2B) systems log the exchange of messages with other parties. Call center packages but also general-purpose CRM systems log interactions with customers. These examples show that many systems have some kind of *event log* often referred to as "history", "audit trail", "transaction log", etc. [1, 2]. The event log typically contains information about events referring to an *activity* and a *case*. The case (also named process instance) is the "thing" which is being handled, e.g., a customer order, a job application, an insurance claim, a building permit, etc. The activity (also named task, operation, action, or work-item) is some operation on the case. Typically, events have a *timestamp* indicating the time of occurrence. Moreover, when people are involved, event logs will typically contain information on the person executing or initiating the event, i.e., the *originator*. Based on this information several tools and techniques for process mining have been developed.[1]

Process mining is useful for at least two reasons. First of all, it could be used as a tool to find out how people and/or procedures really work. Consider for example processes supported by an ERP system like SAP (e.g., a procurement process). Such a system logs all transactions but in many cases does not enforce a specific way of working. In such an environment, process mining could be used to gain insight in the actual process. Another example would be the flow of patients in a hospital. Note that in such an environment all activities are logged but information about the underlying process is typically missing. In this context it is important to stress that management information systems provide information about key performance indicators like resource utilization, flow times, and service levels but *not* about the underlying business processes (e.g., causal relations, ordering of activities, etc.). Second, process mining could be used for *Delta analysis*, i.e., comparing the actual process with some predefined process. Note that in many situations there is a descriptive or prescriptive process model. Such a model specifies how people and organizations are assumed/expected to work. By comparing the descriptive or prescriptive process model with the discovered model, discrepancies between both can be detected and used to improve the process. Consider for example the so-called reference models in the context of SAP. These models describe how the system should be used. Using process mining it is possible to verify whether this is the case. In fact, process mining could also be used to compare different departments/organizations using the same ERP system.

The remainder of this paper is organized as follows. Section 2 introduces the concept of business process mining followed by a short description of a case study in Section 3. Section 4 briefly discusses process mining as a tool for Delta analysis, i.e., measuring the business alignment by comparing event logs with descriptive or prescriptive (process) models.

2 Business Process Mining: An overview

The goal of process mining is to extract information about processes from transaction logs [1]. We assume that it is possible to record events such that (i) each event refers to an *activity* (i.e., a well-defined step in the process), (ii) each event refers to a *case* (i.e., a process instance), (iii) each event can have a *performer*

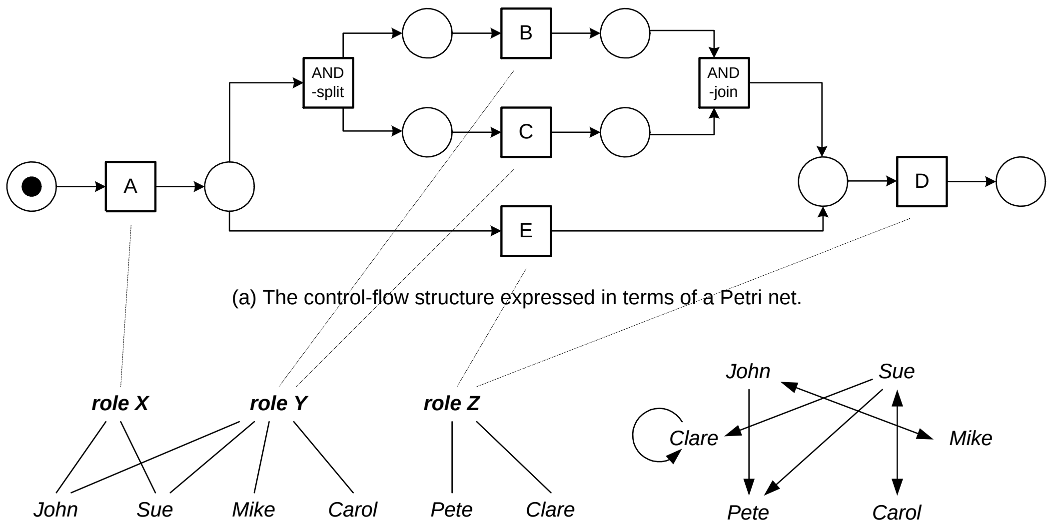
[1] In this position paper, we do not provide an overview of related work, instead we refer to [1, 2] and www.processmining.com.

also referred to as *originator* (the person executing or initiating the activity), and (iv) events have a *timestamp* and are totally ordered. Table 1 shows an example of a log involving 19 events, 5 activities, and 6 originators. In addition to the information shown in this table, some event logs contain more information on the case itself, i.e., data elements referring to properties of the case. For example, the case handling systems FLOWer logs every modification of some data element.

case id	activity id	originator	timestamp
case 1	activity A	John	9-3-2004:15.01
case 2	activity A	John	9-3-2004:15.12
case 3	activity A	Sue	9-3-2004:16.03
case 3	activity B	Carol	9-3-2004:16.07
case 1	activity B	Mike	9-3-2004:18.25
case 1	activity C	John	10-3-2004:9.23
case 2	activity C	Mike	10-3-2004:10.34
case 4	activity A	Sue	10-3-2004:10.35
case 2	activity B	John	10-3-2004:12.34
case 2	activity D	Pete	10-3-2004:12.50
case 5	activity A	Sue	10-3-2004:13.05
case 4	activity C	Carol	11-3-2004:10.12
case 1	activity D	Pete	11-3-2004:10.14
case 3	activity C	Sue	11-3-2004:10.44
case 3	activity D	Pete	11-3-2004:11.03
case 4	activity B	Sue	11-3-2004:11.18
case 5	activity E	Clare	11-3-2004:12.22
case 5	activity D	Clare	11-3-2004:14.34
case 4	activity D	Pete	11-3-2004:15.56

Table 1. An event log.

Event logs such as the one shown in Table 1 are used as the starting point for mining. We distinguish three different perspectives: (1) the process perspective, (2) the organizational perspective and (3) the case perspective. The *process perspective* focuses on the control-flow, i.e., the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths, e.g., expressed in terms of a Petri net or Event-driven Process Chain (EPC). The *organizational perspective* focuses on the originator field, i.e., which performers are involved and how are they related. The goal is to either structure the organization by classifying people in terms of roles and organizational units or to show relation between individual performers (i.e., build a social network). The *case perspective* focuses on properties of cases. Cases can be characterized by their path in the process or by the originators working on a case. However, cases can also be characterized by the values of the corresponding data elements. For example, if a case represent a replenishment order it is interesting to know the supplier or the number of products ordered.



(a) The control-flow structure expressed in terms of a Petri net.

(b) The organizational structure expressed in terms of a activity-role-performer diagram.

(c) A sociogram based on transfer of work.

Fig. 1. Some mining results for the process perspective (a) and organizational (b and c) perspective based on the event log shown in Table 1.

The process perspective is concerned with the "How?" question, the organizational perspective is concerned with the "Who?" question, and the case perspective is concerned with the "What?" question. To illustrate the first two consider Figure 1. The log shown in Table 1 contains information about five cases (i.e., process instances). The log shows that for four cases (1, 2, 3, and 4) the activities A, B, C, and D have been executed. For the fifth case only three activities are executed: activities A, E, and D. Each case starts with the execution of A and ends with the execution of D. If activity B is executed, then also activity C is executed. However, for some cases activity C is executed before activity B. Based on the information shown in Table 1 and by making some assumptions about the completeness of the log (i.e., assuming that the cases are representative and a sufficient large subset of possible behaviors is observed), we can deduce the process model shown in Figure 1(a). The model is represented in terms of a Petri net.

Figure 1(a) does not show any information about the organization, i.e., it does not use any information on the people executing activities. However, Table 1 shows information about the performers. For example, we can deduce that activity A is executed by either John or Sue, activity B is executed by John, Sue, Mike or Carol, C is executed by John, Sue, Mike or Carol, D is executed by Pete or Clare, and E is executed by Clare. We could indicate this information in Figure 1(a). The information could also be used to "guess" or "discover" organizational structures. For example, a guess could be that there are three roles: X, Y, and Z. For the execution of A role X is required and John and Sue have this role. For the execution of B and C role Y is required and John, Sue, Mike and Carol have this role. For the execution of D and E role Z is required and Pete and Clare have this role. For five cases these choices may seem arbitrary but for larger data sets such inferences capture the dominant roles in an organization. The resulting "activity-role-performer diagram" is shown in Figure 1(b). The three "discovered" roles link activities to performers. Figure 1(c) shows another view on the organization based on the transfer of work from one individual to another, i.e., not focus on the relation between the process and individuals but on relations among individuals (or groups of individuals). Consider for example Table 1. Although Carol and Mike can execute the same activities (B and C), Mike is always working with John (cases 1 and 2) and Carol is always working with Sue (cases 3 and 4). Probably Carol and Mike have the same role but based on the small sample shown in Table 1 it seems that John is not working with Carol and Sue is not working with Carol. These examples show that the event log can be used to derive relations between performers of activities, thus resulting in a sociogram. For example, it is possible to generate a sociogram based on the transfers of work from one individual to another as is shown in Figure 1(c).

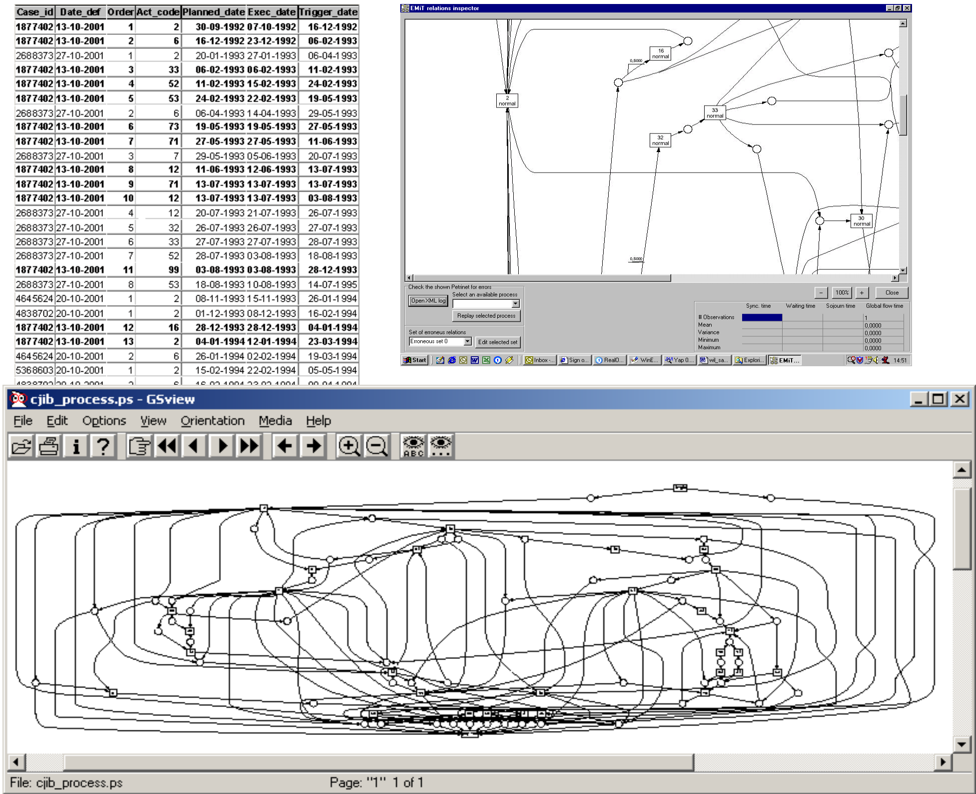
Besides the "How?" and "Who?" question (i.e., the process and organization perspectives), there is the case perspective that is concerned with the "What?" question. Figure 1 does not address this. In fact, focusing on the case perspective is most interesting when also data elements are logged but these are not listed in Table 1. The case perspective looks at the case as a whole and tries to establish relations between the various properties of a case. Note that some of the properties may refer to the activities being executed, the performers working on the case, and the values of various data elements linked to the case. Using clustering algorithms it would for example be possible two show a positive correlation between the the size of an order or its handling time and the involvement of specific people.

Orthogonal to the three perspectives (process, organization, and case), the result of a mining effort may refer to *logical* issues and/or *performance* issues. For example, process mining can focus on the logical structure of the process model (e.g., the Petri net shown in Figure 1(a)) or on performance issues such as flow time. For mining the organizational perspectives, the emphasis can be on the roles or the social network (cf. Figure 1(b) and (c)) or on the utilization of performers or execution frequencies.

To address the three perspectives and the logical and performance issues we have developed a set of tools including (e.g., EMiT , Thumb , and MinSoN) sharing a common XML format (see http://www.processmining.org for more details).

3 Case study

We have applied our mining techniques in several organizations. In this section, we briefly show some results for one of these organizations, i.e., the processes of



Case_id	Date_def	Order	Act_code	Planned_date	Exec_date	Trigger_date
1877402	**13-10-2001**	**1**	**2**	**30-09-1992**	**07-10-1992**	**16-12-1992**
1877402	**13-10-2001**	**2**	**6**	**16-12-1992**	**23-12-1992**	**06-02-1993**
2688373	27-10-2001	1	2	20-01-1993	27-01-1993	06-04-1993
1877402	**13-10-2001**	**3**	**33**	**06-02-1993**	**06-02-1993**	**11-02-1993**
1877402	**13-10-2001**	**4**	**52**	**11-02-1993**	**15-02-1993**	**24-02-1993**
1877402	**13-10-2001**	**5**	**53**	**24-02-1993**	**22-02-1993**	**19-05-1993**
2688373	27-10-2001	2	6	06-04-1993	14-04-1993	29-05-1993
1877402	**13-10-2001**	**6**	**73**	**19-05-1993**	**19-05-1993**	**27-05-1993**
1877402	**13-10-2001**	**7**	**71**	**27-05-1993**	**27-05-1993**	**11-06-1993**
2688373	27-10-2001	3	7	29-05-1993	05-06-1993	20-07-1993
1877402	**13-10-2001**	**8**	**12**	**11-06-1993**	**12-06-1993**	**13-07-1993**
1877402	**13-10-2001**	**9**	**71**	**13-07-1993**	**13-07-1993**	**13-07-1993**
1877402	**13-10-2001**	**10**	**12**	**13-07-1993**	**13-07-1993**	**03-08-1993**
2688373	27-10-2001	4	12	20-07-1993	21-07-1993	26-07-1993
2688373	27-10-2001	5	32	26-07-1993	26-07-1993	27-07-1993
2688373	27-10-2001	6	33	27-07-1993	27-07-1993	28-07-1993
2688373	27-10-2001	7	52	28-07-1993	03-08-1993	18-08-1993
1877402	**13-10-2001**	**11**	**99**	**03-08-1993**	**03-08-1993**	**28-12-1993**
2688373	27-10-2001	8	53	18-08-1993	10-08-1993	14-07-1995
4645624	20-10-2001	1	2	08-11-1993	15-11-1993	26-01-1994
4838702	20-10-2001	1	2	01-12-1993	08-12-1993	16-02-1994
1877402	**13-10-2001**	**12**	**16**	**28-12-1993**	**28-12-1993**	**04-01-1994**
1877402	**13-10-2001**	**13**	**2**	**04-01-1994**	**12-01-1994**	**23-03-1994**
4645624	20-10-2001	2	6	26-01-1994	02-02-1994	19-03-1994
5368603	20-10-2001	1	2	15-02-1994	22-02-1994	05-05-1994

Fig. 2. A fragment of the log of a Dutch governmental institution responsible for fine-collection and the corresponding process mining result.

a Dutch governmental institution responsible for fine-collection.[2] A case (process instance) is a fine that has to be paid. There may be more fines related with the same person. However, each fine corresponds to an independent case. This process has the particularity that as soon as the fine is paid, the process stops. In total there are 99 distinct activities which can be either manually or automatically executed. We selected the fines information for 130136 cases. We constructed the process log and we applied to this log our process discovery method that can handle noisy data [3, 4].

Figure 2 (top-left) shows a fragment of the log containing 130136 cases. This log is generated by an information system specifically constructed for the Dutch governmental institution. (The institution has been in the process of using standard workflow technology but this process has been put “on hold”.) The top-right screen shows a screenshot of our mining tool EMiT while analyzing the log. The bottom screenshot shows the whole process obtained through application

[2] The name of the organization is not given for reasons of confidentiality. We want to thank L. Maruster, R. Dorenbos, H.J. de Vries, H. Reijers, and A. in t Veld for their valuable support.

of the process mining techniques. The discovered models have been inspected by the domain experts. They concluded that our discovered models were able to grasp the important aspects of the process. Moreover, the discovered models revealed aspects that are often questioned when discussing the process model. These experiences showed that process discovery can provide useful insights into the current practice of a process and highlight difference between the actual process and the prescriptive/descriptive model [4].

We have also applied our process mining techniques to a health-care process where the flow of multi-disciplinary patients is analyzed. We have analyzed event logs (visits to different specialists) of patients with peripheral arterial vascular diseases of the Elizabeth Hospital in Tilburg and the Academic Hospital in Maastricht. Patients with peripheral arterial vascular diseases are a typical example of multi-disciplinary patients. We have preliminary results showing that process mining is very difficult given the "spaghetti-like" nature of this process. Only by focusing on specific tasks and abstracting from infrequent tasks we are able to successfully mine such processes. Given this experience we are now focussing on processes have more structure. For example, environments using case handling system like FLOWer (the workflow product of Pallas Athena), e.g., the Employee Insurance Implementing Body (Uitvoering Werknemersverzekeringen, or UVW).

4 Delta analysis

Process mining can be used for *Delta analysis*, i.e., comparing the actual process with some predefined process representing the information system. Note that in many situations there is a *descriptive* or *prescriptive* process model. Such a model specifies how people and organizations are assumed/expected to work. By comparing the descriptive or prescriptive process model with the discovered model, discrepancies between both can be detected and used to improve the process.

As an example consider the so-called reference models in the context of SAP. These models, typically expressed in terms of Event-driven Process Chains (EPC), describe how the system should be used. However, in practise people may use the system differently or only use a subset of the system.

Consider for example the two EPCs shown in Figure 3. The left EPC represents the full process, the right EPC shows the parts that are actually being used. Note that tools like the SAP Reverse Business Engineer (RBE) can be used to monitor how frequent parts of the SAP system are being used. Unfortunately, tools like RBE do not consider the order of activities nor other aspects such as the organizational and case perspective.

For the Fifth Workshop on Business Process Modeling, Development, and Support (BPMDS'04) with the theme "Creating and maintaining the fit between business processes and support systems", we would like to discuss the following statement: "Process mining can be used to improve the fit between business processes and information systems". In other words: A Delta analysis based on

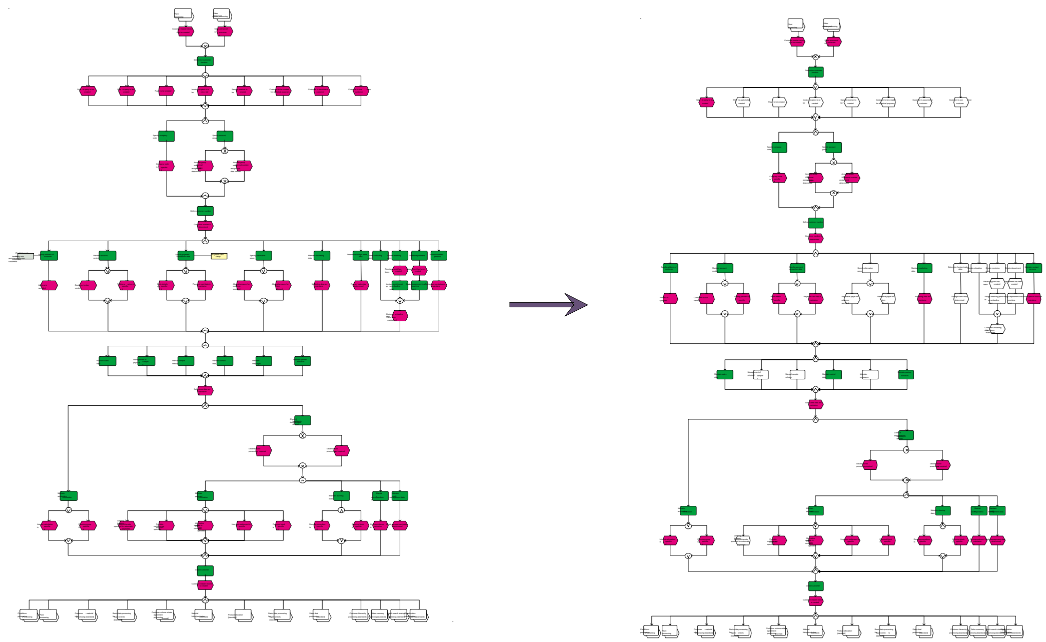


Fig. 3. Two EPCs: the full EPC (left) and EPC really being used (right).

the actual event logs and some descriptive or prescriptive of the process and its organization supports business alignment.

Acknowledgements
The author would like to thank Ton Weijters, Boudewijn van Dongen, Ana Karla Alves de Medeiros, Minseok Song, Laura Maruster, Eric Verbeek, Monique Jansen-Vullers, Hajo Reijers, Michael Rosemann, and Peter van den Brand for their on-going work on process mining techniques. Parts of this position paper have been based on earlier papers with these researchers.

References

1. W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A.J.M.M. Weijters. Workflow Mining: A Survey of Issues and Approaches. *Data and Knowledge Engineering*, 47(2):237–267, 2003.
2. W.M.P. van der Aalst and A.J.M.M. Weijters, editors. *Process Mining*, Special Issue of Computers in Industry, Volume 53, Number 3. Elsevier Science Publishers, Amsterdam, 2004.
3. W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster. Workflow Mining: Discovering Process Models from Event Logs. QUT Technical report, FIT-TR-2003-03, Queensland University of Technology, Brisbane, 2003. (Accepted for publication in IEEE Transactions on Knowledge and Data Engineering.).
4. L. Maruster. *A Machine Learning Approach to Understand Business Processes*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2003.